Against the Grain

Volume 15 | Issue 6 Article 11

December 2003

Getting the Best Out of It! Usage Analysis from the Publishing Perspective

Marthyn G.M. Borghuis
ScienceDirect, M.Borghuis@elsevier.com

Recommended Citation
Borghuis, Marthyn G.M. (2003) "Getting the Best Out of It! Usage Analysis from the Publishing Perspective," *Against the Grain*: Vol. 15: Iss. 6, Article 11.
DOI: https://doi.org/10.7771/2380-176X.4202

Usage Statistics at Yale ...
from page 32

send them every quarter and do not. When I do receive them, the electronic journal statistics contain one number: downloads per journal per month. While this is the essential number I want to know, I get no sense how many tables of contents were viewed, whether the package as a whole was searched, much less any linking information. But judging from my cursory survey above, the situation is improving, though still time consuming. I also find most vendors to be receptive to complaints or questions about their statistics and have answered at least five surveys in the last calendar year about electronic resources that had some component devoted to usage statistics.

The inability of even some expensive vendors to provide us with the minimum level of acceptable statistics leads to holes in our analysis of how people are using electronic resources. **Yale** has been a member of the **ARL E-Metrics** project for the last few years, and as part of this we are asked to test various statistical measures of electronic usage and collections in libraries. The most troublesome question is "Total Number of Searches on Electronic Databases." While it is possible (though incredibly time consuming) for me to give a number of "total searches we are told about," it is unclear what this can really mean to us. It is difficult to extrapolate from our available statistics to what our actual number of database searches were. We also can't tell how different groups are using our electronic resources, or how people are moving from one resource to another, or whether they use our linking tools or are unaware of them. This is important as we assist users in person and implement new services to exploit our full electronic potential.

My future goals for our usage statistics are to increase their relevance to librarians at **Yale**. First, I continue to strive to get information that can really help them make decisions about what kinds of resources are the best for our users. I also want to be able to demonstrate trends across subject areas, not just vendors. Setting up a "data farm" or database of the statistics is one way we can accomplish this. Finally, we support efforts to get better statistics from vendors and contribute to new research using this data, at **Yale**, within our consortia, and at the national level.

Our collection of usage statistics, while perhaps not as comprehensive as we'd ideally like it to be, is a step in the direction of being able to tell what our users are doing and what they want and need, even if they can't verbalize it to us. Sitting in our library, it can be hard to tell if the investment **Yale** made in electronic books was worth the money. Once we created an eBooks database and put it prominently on our Web page, usage of electronic books increased immediately. Certainly it is hard to tell how satisfied people are with those books aside from the fact that usage keeps growing every month. But we do know that people are using them, and perhaps these statistics and our experiences in collecting them will lead to getting a more complete picture of our electronic resource use in the future.

Getting the Best Out of It! Usage Analysis from the Publishing Perspective

by **Marthyn G.M. Borghuis** (ScienceDirect, Senior Manager, Usage Research, Amsterdam, The Netherlands) <M.Borghuis@elsevier.com>

What Factors Drive Usage?

The conversion from print to electronic publishing is proceeding rapidly as is the retrospective digitisation of the archive of print articles. Availability of full text articles in electronic format does have a substantial impact on the volume of usage data generated, this is obvious. It seems, however, that for publishers the three crucial factors influencing usage levels are:

1. The number of universities, corporations, hospitals, and research organizations worldwide that have access to the online journal collection,
2. The number of users, within these institutes that have access to the collection, and finally,
3. The increase in the number of readings (and thus usage) by these users (usage intensity).

The Global Trend

Knowing these main usage factors, what then is the global usage trend across all publishers offering primary content online? Where is the sky and does it have limits? It is an opportunity to present to the readers of ***Against the Grain***, a straightforward analysis of online journal usage developments worldwide to date and in the years to come. For that exercise the usage data of **ScienceDirect** have been used covering the period Jan 2001 to Sep 2003.

Analysing **Elsevier's ScienceDirect** data, the following assumptions can be made: Extrapolating from **ScienceDirect** usage and the size of its journal collection relative to the total scientific journal literature available online, it is likely that across all primary publishers at least 90 million requests for full text articles were made on a global scale in September 2003. In addition it is essential to know, that full-text usage is increasing by an average 15% in peak months (September to November & February through April) and about 8% in all other months. As a consequence, every year so far, usage has more than doubled and there is no sign whatsoever that this will decline in the next two to three years! It is thus expected that total full text article usage achieved in 12 months, across all libraries and publishers worldwide, will pass the **1 billion landmark** in early 2004!

Looking at this trend it is very clear, that the transfer from library print collections to electronic full text collections has resulted in a tremendous increase in accessibility and thus usage of full text scholarly and scientific articles.

In itself this is a true success story!

What is Contributed by each of the Usage Drivers?

a. Number of Institutes and Users having access:

A **King & Tenopir** (2002) analysis resulted in a count of 6.8 million scientists in the USA "engaged in research and development and teaching, who were likely to be readers of scientific publications."[1] Numbers are based both on professional scientists in academia, corporations, government, and research organizations. Knowing these US numbers, how large will the potential reader base be on a global level? There is evidence that the US portion is around 30% and thus that the worldwide number of scientists, R&D staff, and educators is approximately 20 million. After starting in 1999, the present size of the **ScienceDirect** customer base is 70% of the total estimated number of research institutes (i.e., institutes that maintain a library facility, which serves users, who read from scientific and scholarly journals). As a library has a collection from various publishers, other primary publishers will also have reached those institutes. There might be a slight variance in the total numbers for each publisher, because publishers with a small and narrow collection will have a lower number of potential customers.

b. Usage-intensity:

In their earlier study **King** and **Tenopir** found that readings in the non-academic sector amount to 102 per researcher per annum, and in academia to 188.[2] However, the authors report also that the number of scientists residing outside academia is a lot higher than that within academia. It is thus necessary to set the average readings at a lower level and estimate 135 readings per annum across all readers. The authors recently updated their findings and report a continuing increase in average annual readings per academic reader, up to 216 for 2003.[3]

Adapting this number again for scientists outside academia reading less, the average number of annual readings across all US scientists should be set at 153. Finally, the authors take another factor into account: about 60% of all readings are still from print, which leaves 40% from electronic sources. The equation is complete when estimating that about 70% of all institutes do have access to online articles. This means that total annual US readings from electronic sources can be found by the following equation: 153*6,8M*0.40*0.7 = 290M. This result is pretty close to the expected US 2003

continued on page 36

Getting the Best Out of It!
from page 34

usage market share which is 0.3*1.0 billion = 300M. The difference between the two totals can be explained by the assumption that not all articles retrieved electronically are read, due to crawler or robot activity among other reasons.

The authors further report that searching online bibliographic databases on one hand, and using enlarged electronic library collections on the other, are the two major factors that caused the increase in readings per academic reader and thus usage per user. In addition, they report a significant increase in readings from library collections and thus a diminishing share of readings from personal and member subscriptions. They also confirm more readings of electronic titles not previously subscribed to in print.

Collection Management: Libraries and Publishers Alike

Once publishers started to publish scholarly and scientific journals online and to collect and analyze usage of these journals, dramatic changes occurred in the way librarians and publishers communicate. The contact between the two parties is now direct, which immediately generates a need to better understand each other's business processes. The reporting of online usage of full text articles is a new phenomenon. At the time when only print collections existed, it was a cumbersome and often difficult exercise for either librarians or primary publishers to quantitatively establish the usage of journal collections. For librarians, available lending data (including interlibrary loan balances) per journal title was used to measure the attractiveness of a journal collection. For unbound issues, techniques like reshelving counts were applied; these techniques proved to be both time-consuming and often incomplete.

Publishers faced the same difficulties: measuring subscription figures, subscription income, or the number of articles or printed pages published, proved to be an inaccurate measure of the attractiveness of published journals. Now, for the first time in history, both librarians and publishers are able to share the same usage information and, what is more, both parties should now be better equipped to have a better-informed opinion on the relevance and attractiveness of the e-journal collections they subscribe to or produce.

What Do Publishers Do with Usage Data?

First of all, I would advise publishers to make their own usage reports **COUNTER** compliant (for details on **COUNTER** see the contribution of **Peter Shepherd** in this issue of ***ATG***) and not only the reports for their customers, but also those for in-house evaluation of their journal collections. The main drivers in primary publishing, however, are the editorial boards: they define the aims and scope of a journal and organize the refereeing process and peer review, thus guaranteeing the uniqueness of published research results and preventing duplication. It is a basic requirement to maintain a certain quality standard for articles to be published, which will enable librarians to earmark these publications.

A publisher's journal management staff, however, may add a new criterion to the ones they already use to evaluate the composition of the journal portfolio for which they are responsible. Circulation figures are supplemented by data about the number of (repeat) users and usage intensity of a journal. Usage analysis will also help in sustaining new and quickly developing areas of research and at a far earlier stage than other measures, such as citation data, can tell.

Comparing usage and citations from another perspective could reveal that there are substantial subject-specific differences in the volume of usage and citations. Especially in areas like health, clinical medicine, engineering, social sciences or psychology, usage outperforms the relative level of citations quite substantially. These areas do attract a lot of users, who seem to be active in more fundamental science areas; they do want to combine their research with the discoveries in these other subject fields. Many users do also read articles in these areas out of personal interest, while working in other faculties of science. In this sense both groups are the "virtual" representation of so-called "walk-in" users as known in the print based libraries. There are also areas where reading has always been predominant and where writing and publishing articles is less common. Here we can localize the users active in the various engineering departments.

Finally, more detailed studies point out that for publishers like **Elsevier**, with a broad collection of journals, review book-series and major reference works *in electronic format*, it is obvious that not all science areas benefit from the strong increase in usage to the same degree. High usage areas are definitely found in the life sciences, chemistry, and environmental sciences.

To end with a controversial statement: where institutional print subscriptions see a rapid conversion to the electronic-only versions open to and accessible by the community of readers, it might occur that print based, personal and member subscriptions will only gradually disappear as they incorporate and bundle quite a few excellent reader needs at once: alerting (in the postbox or on the doormat), print-at-hand, and no need to log-in.

Endnotes

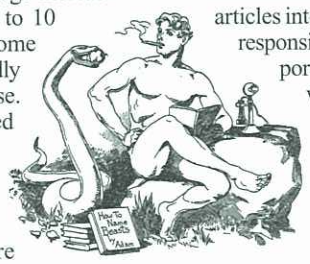
1. Tenopir, Carol and King, Donald W. *Towards Electronic Journals, Realities for Scientists, Librarians and Publishers,* Special Libraries Association, 2002, p.18-19.
2. See Tenopir and King (2002) p.19.
3. King, et. al., found a continuing increase in the number of readings per reader especially since the electronic availability of articles. See: King, Donald W., Tenopir, Carol, Hansen Montgomery, Carol Aerni, Sarah E. Patterns of Journal Use by Faculty at Three Diverse Universities, *D-Lib Magazine 2003 (Oct) http://www.dlib.org/dlib/october03/king/10king.html.*

The Beginning of Value Assessment: Usage Information in the E-Journal Age

by **John Sack** (Director, HighWire Press, Stanford University; Phone: 650-723-0192) <sack@stanford.edu>

Right from the start of **HighWire Press** with the publication of the *Journal of Biological Chemistry* online in 1995, there was a strong interest in learning what we could from usage statistics, or "server logs." We knew we should measure just about anything that moved or clicked, because the early days of e-journals were like the early days of space exploration: we didn't know what observations would be useful because we didn't know what new phenomena in user behavior would occur. With over 350 online journals, from over 150 different societies and publishers, about all the HighWire-hosted publishers could agree on was that we needed to measure "hits."

At first, each journal editor wanted to watch the hits climb - and they did climb: a number of journals in the 1990's saw usage increase by factors anywhere from 2 to 10 from one year to the next as some whole user communities rapidly shifted from print to online use. But soon after, editors realized that they had data on use patterns that they couldn't really obtain from counting print copies mailed, or even by reader surveys: what are readers reading?



This began our understanding that electronic journal usage information could be "mined" for much more than hits: it could tell those who have intellectual responsibility for collecting articles into journals - editors - and those with responsibility for collecting journals to support academic programs - librarians — what role a journal had in the intellectual life of a discipline or institutional community.

We also realized that the initial introduction of a new technology such as e-journals was the best and perhaps only time to instrument the shift in literature-research pro-

continued on page 38